HAL
open science

A Non-symmetrical Method of Image Local-Difference Comparison for Ancient Impressions Dating

Etienne Baudrier, Nathalie Girard, Jean-Marc Ogier

To cite this version:

Etienne Baudrier, Nathalie Girard, Jean-Marc Ogier. A Non-symmetrical Method of Image Local-Difference Comparison for Ancient Impressions Dating. Liu Wenyin, Llados Josep, Ogier Jean-Marc. Graphics Recognition. Recent Advances and New Opportunities 7th International Workshop, GREC 2007, Curitiba, Brazil, September 20-21, 2007. Selected Papers), Springer Berlin / Heidelberg, pp.257-265, 2008, Lecture Notes in Computer Science, 10.1007/978-3-540-88188-9_24 . hal-00385323

HAL Id: hal-00385323

https://hal.science/hal-00385323v1

Submitted on 19 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Non-Symmetrical Method of Image Local-Difference Comparison for Ancient Impressions Dating

Étienne Baudrier[1], Nathalie Girard[2], and Jean-Marc Ogier[2]

[1] Laboratoire Signal Images Communication
Université de Poitiers, Bvd Marie et Pierre Curie, BP 30179, 86962 Futuroscope Chasseneuil Cedex, France
etienne.baudrier@sic.univ-poitiers.fr
[2] Laboratoire d'Informatique, Image et Interactions
Université de La Rochelle, Avenue Crépeau, 17042 La Rochelle Cedex 1, France
nathalie.girard@etudiant.univ-lr.fr,jean-marc.ogier@univ-lr.fr

Keywords

Image comparison, binary images, Hausdorff distance, local dissimilarity measure, visualization, ancient images, dating

Abstract. In this article, we focus on the dating of images (impressions, ornamental letters) printed starting from the same stamp. This difficult task needs a good observation of the differences between the compared images. We present a method, based on a local adaptation of the Hausdorff distance, that evaluates locally the image differences. It allows the user to visualize these differences. A description of the pertinent differences for the dating allows us to evaluate our method visualization ability. Then our method is successfully compared to the existing method. Finally, a framework for a future automatic dating method is presented.

1 Introduction

As ancient documents are being digitized, systems for retrieving documents or images can now be found in Digital Libraries [1]. With regard to illustrations, the content-based image retrieval is difficult and the user often needs to check visually the similarity of the retrieved images. In this article, we focus on the dating of images (impressions, ornamental letters) printed starting from the same stamp. This issue is difficult even for the expert eyes. In a perfect case where printings are perfectly conserved, digitized and registered, the sign of the map of pixel-to-pixel gray level difference (PPDMap) would be sufficient to conclude: if the difference is positive, it means that the second printing has been printed with the stamp in a degraded state, and so the second printing is less old. Nevertheless, in a real case, printings starting from the same stamp can differ for various reasons:

1. The printing degradation state,
2. The digitalization which can cause variations in the gray level, the resolution, etc,
3. If there is a binarization step, the method used can cause differences in the binarized images,
4. The registration can cause a slight shift or/and rotation resulting in differences in the visualization,
5. The differences due to the wood stamp ageing.

The only interesting differences for the user are the ones due to the stamp degradation. We call the other ones *perturbations*.

In this frame, we present a method that can help the expert's dating. This method, based on a local adaptation of the Hausdorff distance, evaluates locally the differences. It minimizes the perturbation impact and allows a better visualization of the interesting differences. No other visualization method exists than the PPDMap. Our method performance has been successfully compared to the PPDMap in [2]. Finally, the frame of an automatic dating method is detailed.

2 Dissimilarity Measure Based on the Hausdorff Distance

Among dissimilarity measures over binary images, the Hausdorff distance (HD) has often been used in the content-based retrieval domain and is known to have successful applications in object matching [3] or in face recognition [4]. For finite sets of points, the HD can be defined as [3]:

Definition 1 (Hausdorff distance). *Given two non-empty finite sets of points $F = (f_1, \ldots, f_n)$ and $G = (g_1, \ldots, g_m)$ of $\mathbb{R}^2$, and an underlying distance d, the HD is given by*

$$HD(F, G) = \max\left(h(F, G), h(G, F)\right) \tag{1}$$

where

$$h(F, G) = \max_{f \in F}\left(\min_{g \in G} d(f, g)\right). \tag{2}$$

$h(F, G)$ is the so-called *directed Hausdorff distance*.

The classical HD presents interesting properties but measures the most mismatched points between F and G, and presents as main drawback its sensitivity to noise [5]. Indeed, considering two images containing the same pattern and one point added to the first image, far from the pattern, then the HD will measure the distance between the pattern and the point.

Several modifications of the HD have been proposed to improve it such as: the partial HD [3], the modified HD (MHD) [6], the censored HD [5], the "doubly" modified Hausdorff distance [4], the least trimmed squared HD [7] and the

weighted Hausdorff distance [8]. Those improved HD are detailed in [9]. These measures stay global and do not take into account local dissimilarities. Indeed, if $D_H(F,G) = \alpha$, it means that there are $f \in F, g \in G$ that realize the maximum and the minimum in Eq. (1): $d(f,g) = \alpha$. But the measure $D_H(F,G) = \alpha$ does not allow one to say if the couple (f,g) is unique or if there are several couples of points realizing the distance α and, in this case, if the couples are gathered in a part of the images or distributed everywhere in the image, which corresponds to different degrees of dissimilarity. These observations motivated us to define a local and parameter-free HD which in the next paragraph.

2.1 Definition of the Windowed Hausdorff Distance

The main reasons of the modification is that the DH is not defined for empty sets and this case is possible in a window. Moreover, the obtained measures when the window is sliding or growing must be consistent. A solution is to introduce the distance to the window side as it follows:

Definition 2 (Windowed Hausdorff distance). *Let F, G be two bounded sets of $\mathbb{R}^2$.*
$HD_W(F,G) = \max\left(h_W(F,G), h_W(G,F)\right)$
where there are three cases

1. *If $F \cap W \neq \emptyset$ and $G \cap W \neq \emptyset$,*
$$h_W(F,G) = \max_{f \in F \cap W} \left[\min \left(\min_{g \in G \cap W} d(f,g), \min_{w \in Fr(W)} d(f,w) \right) \right],$$
2. *if $F \cap W \neq \emptyset$ and $G \cap W = \emptyset$,*
$$h_W(F,G) = \max_{f \in F \cap W} \left[\min_{w \in Fr(W)} d(f,w) \right],$$
3. *if $F \cap W = \emptyset$,*
$$h_W(F,G) = 0.$$

remark 1 – *In case both of the sets are non-empty, the only difference with the classical definition is the term $\min_{w \in Fr(W)} d(f,w)$ which is the distance from the point f to the edge.*
– *In case there is exactly one set without point in W, one of the two directed distances is equal to 0 and the expression of the other one takes into account the distance to the edge.*
– *In case there is no point of F or of G in W, both of the directed distances are equal to 0 and therefore the global distance too. This is consistent with the fact that the two extracted parts are equal.*

The definition of the windowed HD enables to make a local distance but it introduces a parameter which is the window size. It can be chosen by the user, or automatically and globally, or locally according to the local surrounding. The following properties of the windowed HD allow to fix locally the window size and then to evaluate the local dissimilarity.

property 1 (Identity) *Let F, G be two bounded sets of points of $\mathbb{R}^2$, and W a convex closed subset of $\mathbb{R}^2$.*

$$HD_W(F,G) = 0 \Longleftrightarrow F \cap W = B \cap W \tag{3}$$

The following properties need the window W to be a ball. Prop. 2 ensures that the new pieces of information that are taken into account when the window is enlarged do not reduce the former dissimilarity-measure value. Prop. 3 gives a maximum to the windowed HD.

property 2 (growth) *Let $V = B(x_v, r_v)$ and $W = B(x_w, r_w)$ be two close discs such as $V \subset W$ then $HD_V(F,G) \leq HD_W(F,G)$.*

property 3 (Boundary) *Let $x \in \mathbb{R}^2$ and $r > 0$, and let define $W = B(x,r)$ then $HD_W(F,G) \leq HD(F,G)$.*

An algorithm for the computation of the local HD map is proposed below (alg. 1). It consists of a sliding window whose radius is locally adapted to find the local optimal radius.

Algorithm 1 Computation of LDMap

```
compute D_H(F,G)
for all pixel x do
  n := 1 {initialization of the window-size}
  while HD_{B(x,n)}(F,G) = n and n ≤ HD(F,G) do
    n := n + 1
  end while
  LDMap(x) = HD_{B(x,n-1)}(F,G) = n - 1
end for
```

It shows the way to adapt the window to the local dissimilarity. This step is done in the *while* loop.

Nevertheless, this algorithm is time consuming. Indeed, the computation complexity is in $O(m^4)$ for two $m \times m$ pixel images. The next section presents a formula for the measure that saves most of the time computation and gives the same result (detail can be found in [10]). The computation is faster but the interpretation –in terms of local dissimilarity measure– comes from Alg. 1.

2.2 Local Dissimilarity Map

Theorem 1 (LDMap mathematical formula).

$$\forall x \in \mathbb{R}^2,\ LDMap(x) = |G(x) - F(x)| \max(d(x,F), d(x,G))$$

The formula gives for each pixel x a value that depends on the distance transformation from the sets F and G. Fast algorithms have been developed for distance

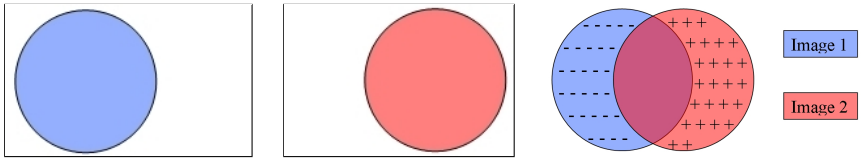


Fig. 1. Asymmetry illustration : Two images and the sign of their SILDMap

transformation. Their computation complexity are $O(m^2)$ for $m \times m$ images. So the LDMap complexity with the formula is a $O(m^2)$, which is linear in the pixel number.

The LDMap provides symmetric differences measures :

$$LDMap(F, G)(x) = LDMap(G, F)(x).$$

In order to date a printing against another, an asymmetry is introduced in the LDMap, in the following way: if a pixel group is present (resp absent) in the former printing and not in the latter, it is negatively (resp positively) measured (see Fig. 1 where the sign in SiLDMap is represented).

The Signed Local Dissimilarity Map (SiLDMap) gathers all the signed measures in a map :

Theorem 2 (SiLDMap).

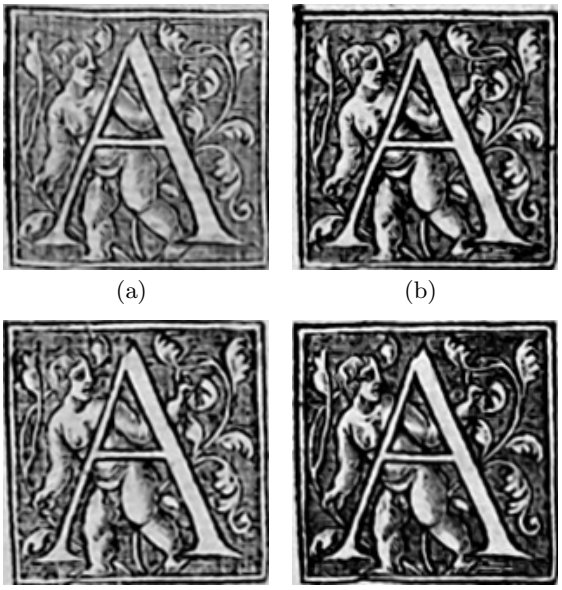
For x, a pixel of the images,

$$SiLDMap(x) = (G(x) - F(x)) \max(d(x, F), d(x, G)).$$

3 Experiment and Perspectives

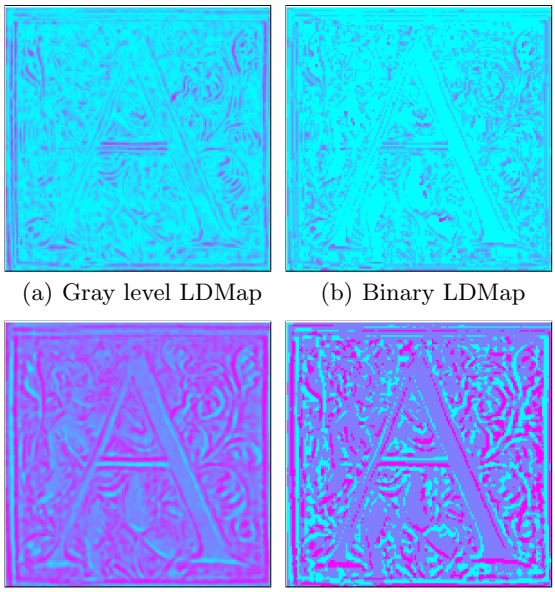
3.1 Experiment

To assess the efficiency of LDMap, let's evaluate first its ability to minimize the impact of perturbations and its ability to render the relevant ones. We have used a database coming from the digital library BVH [11] which includes 168 images of ornamental letters. The database contains four versions of each ornamental letter stamp, coming from four distinct books (so the database contains impressions of 168/4=42 distinct stamps). The four versions of the same stamp that are available provide some perturbations in the visualization: perturbations of ageing, digitization and registration (see Fig. 2). The tested methods are used to produce maps (see Fig. 3) that are classified by a support vector machine (SVM). The experiment protocol is as follows: the comparison of the 168 images gives 14028 visualization maps that are separated in two classes, one gathering the 252 maps comparing images from the same stamp C_{sim} and one including the 13776 maps comparing images from distinct stamps C_{dissim}. A SVM learning stage is done on a part of the two classes and a test is realized on the other part. The classification results are compared with those obtained manually. The



(a) (b) (c) (d)

Fig. 2. A group of four distinct printings coming from the same stamp



(a) Gray level LDMap (b) Binary LDMap

(c) Gray level PPDMap (d) Binary PPDMap

Fig. 3. Visualization Maps between the the ornamental letters 2(a) and 2(c). PPDMaps contain more high values than the LDMap: they are more sensitive to perturbations.

results, reflecting the average found on 100 tests, are gathered in Tab. 1. Precision and recall measures do not bring more information because they are up to 96% since the first item retrieval rate. Results show that the LDMap allows the

Successful retrieval	found in $\mathcal{C}_{sim}$	found in $\mathcal{C}_{dissim}$
Gray level LDMap	**95%**	**97%**
Binary LDMap	93%	95%
Gray level PPDMap	70%	75%
Binary PPDMap	70%	69%

Table 1. Results of the classification for the gray level and binary LDMap and PPDMap.

SVM to make a better classification than the PPDMap. So the perturbations are less represented in the LDMap than in the PPDMap. The LDMap visualization is therefore better than the PPDMap one. One reason is that a PPDMap does not enable the user to distinguish between a simple translation and a real difference. The result is a successful retrieval rate of 96%, which proves the LDMap robustness against perturbations. A study of the LDMap robustness to ink-stain and erasing can be found in [10]. The robustness is really better for a stain than for an erasing. One reason is that treated information is the one of black pixels. As a consequence, the stain does not change so much the LDmap values whereas the erasing produces a great increase of the LDMap values. Nevertheless, for stains and erasing with a surface smaller than 20% of the total image surface, the robustness is really good.

3.2 Visualisation and Discussion

Thus, the SiLDMap, the signed version of the LDMap, can help in the printing dating by quantifying and localizing relevant differences. The easiest way is to use the SILDMap to visualize the printing differences. Fig. 4 gives an example from two printings of an ornamental letter, and their SILDMap. It contains positive and negative values so the dating is not trivial. Four significant differences have been surrounded in blue. The one at the top of the letter "L" is the only negative and is due to a inking difference. The one at the bottom of the "L" can be interpreted as a missing piece of wood in the first printing, which leads to think that the wood stamp was older when it has been used for the first printing than for the second printing.

The SILDMap measures the differences locally so it is sensible to registration. Fig. 5 shows an example of registration issue: the values in the centre of the SILDMap are low which proves that the affine registration is good, nevertheless high values increase toward the bottom right corner and the top left corner. As both of the printing come from the same stamp, it means that one of their digitized images is deformed (which is an interesting piece of information). The

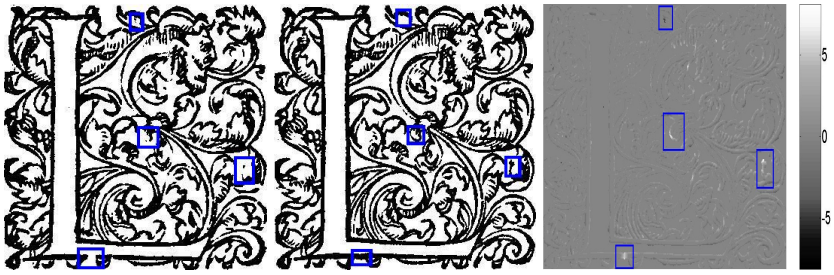


Fig. 4. Two impressions and their SiLDMap

deformation brings high values in the SILDMap that hide the pertinent differences. So a non-linear registration is necessary (and the SILDMap may contain information for the registration) to exploit the SILDMap efficiently.

Fig. 5. Two impressions and their SiLDMap illustrating the perturbation impact

3.3 Perspectives

The next step is to date automatically the printings thanks to their SILDMap. As Fig. 4 shows it, the dating is not only based on the difference values, but also on elements like local connectedness, inking pressure... An automatic diagnosis should then associate high level information with the SiLDMap to be efficient.

References

1. Baird, H.S.: Digital library and document image analysis. In: Proc. of the 7th Int. Conf. on Document Analysis and Recognition (ICDAR), IAPR (2003) 1–13
2. Baudrier, E., Riffaud, A.: A method for image local-difference visualization. In: Proc. of the 9th Int. Conf. on Document Analysis and Recognition (ICDAR), Brazil, IAPR (2007)
3. Huttenlocher, D.P., Klanderman, D., Rucklidge, W.J.: Comparing images using the Hausdorff distance. Trans. on Pattern Analysis and Machine Intel. **15**(9) (1993) 850–863

4. Takàcs, B.: Comparing faces using the modified Hausdorff distance. Pattern Recognition **31**(12) (1998) 1873–1881
5. Paumard, J.: Robust comparison of binary images. Pattern Recognition Letters **18**(10) (1997) 1057–1063
6. Dubuisson, M.P., Jain, A.K.: A modified Hausdorff distance for object matching. In: Proc. of the Int. Conf. on Pattern Recognition (ICPR), IAPR (1994) 566–568
7. Sim, D.G., Kwon, O.K., Park, R.H.: Object matching algorithms using robust Hausdorff distance measures. IEEE Trans. on Image Processing **8**(3) (1999) 425–429
8. Lu, Y., Tan, C., Huang, W., Fan, L.: An approach to word image matching based on weighted Hausdorff distance. In: Proc. 6th Internat. Conf. on Document Anal. Recogn. (2001) 921–925
9. Zhao, C., Shi, W., Deng, Y.: A new Hausdorff distance for image matching. Pattern Recognition Letters (2004)
10. Baudrier, E., Millon, G., Nicolier, F., Ruan, S.: Binary-image comparison with local-dissimilarity quantification. Pattern Recognition **41**(5) (2008) 1461–1478
11. Ramel, J., Busson, S., Demonet, M.: Agora: the interactive document image analysis tool of the BVH project. In: Conf. on Document Image Analysis for Library. (2006) 145–155